PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ^{6} : C12Q 1/68	A1	(11) International Publication Number: **WO 99/58721**
		(43) International Publication Date: 18 November 1999 (18.11.99)

(21) International Application Number: PCT/US99/10417

(22) International Filing Date: 12 May 1999 (12.05.99)

(30) Priority Data:
09/076,575   12 May 1998 (12.05.98)   US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US   09/076,575 (CON)
Filed on   12 May 1998 (12.05.98)

(71) Applicant (for all designated States except US): WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): WANG, David, G. [CN/US]; 6 Hansel Road, Newtown, PA 18940 (US). LANDER, Eric, S. [US/US]; 151 Bishop Allen Drive, Cambridge, MA 02138 (US).

(74) Agents: HOGLE, Doreen, M.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US) et al.

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MULTIPLEX DNA AMPLIFICATION USING CHIMERIC PRIMERS

(57) Abstract

The present invention provides for a method of simultaneously amplifying a large number of target sequences from a template nucleic acid and subsequently labeling the amplification products.

-1-

MULTIPLEX DNA AMPLIFICATION USING CHIMERIC PRIMERS

RELATED APPLICATIONS

This application is a Continuation of U.S. patent application serial number 09/076,575, filed May 12, 5   1998, the entire teachings of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

15   The Human Genome Project and associated research has produced a series of genetic maps of increasing density. Variation, *i.e.*, polymorphism, is the foundation of any genetic map, and the types of polymorphisms used to build genetic maps have evolved 20   over the years. Differences in gross morphology were used to construct the first genetic maps, but presented difficulties in that (1) morphological traits can be difficult to detect reliably, (2) there were few such reliable traits, which meant that very large populations 25   had to be used to detect linkage between traits, and (3) similar morphology can be caused by several different

genes. In the 1970's, isozymes provided additional
variation for constructing genetic maps. Restriction
fragment length polymorphisms (RFLPs) and restriction
site maps supplanted isozymes as the tools of choice in
5   the 1980's. In the 1990's, simple sequence length
polymorphisms (SSLPs, also called STRs, or Short Tandem
Repeats) and sequence tagged sites (STSs) allowed
construction of maps containing greater numbers of
markers than ever before.

10   Single nucleotide polymorphisms (SNPs) are
positions at which two alternative bases occur at
appreciable frequency (>1%) in the human population, and
are the most common type of human genetic variation.
There is growing recognition that large collections of
15 mapped SNPs would provide a powerful tool for human
genetic studies. Because this type of variation is at
the sequence level, it also opens a window on the root
causes of variation, including differences in gross
morphology and biochemistry, and susceptibility to
20 genetic diseases. SNPs can also be used to create more
markers for genetic maps, or to study linkage
disequilibrium, or human evolution and migration.
Before SNPs can be systematically applied in such
studies, however, it is necessary to create a large
25 collection of such loci, construct maps of their genomic
locations, and develop methods for large-scale
genotyping.

There have been many attempts to analyze large
numbers of samples simultaneously, a method often
30 referred to as "multiplexing." Attempts have been made
to formulate such methods for high throughput sequencing
(Church, G.M., and Kieffer-Higgins, S., Science 240:185-
188 (1988)) and thermocycling (Shuber, A.P. et al.,
Genome Res. 5:488-493 (1995); Edwards, M.C., and Gibbs,
35 R.A., PCR Meth. Appl. 3:S65-S75 (1994); Chamberlain,
J.S. et al., Nuc. Acids Res. 16:11141-11156 (1988)), but
they have either met with variable success, or have been

able to multiplex only a few samples at a time. With
the discovery of increasing numbers of SNPs, there
exists a real need to increase the efficiency with which
they can be analyzed.

5   SUMMARY OF THE INVENTION
        The present invention relates to a method of
multiplex amplification by which a large number of
target sequences on a template nucleic acid are
amplified simultaneously, and labeled for detection.
10   During an amplification reaction, the primers are
incorporated into the products. In the early cycles of
the reaction, the primers copy the target sequences on
the template nucleic acid. In later cycles of the
reaction, however, the number of copies outnumbers (and
15   later overwhelms) the number of original template
sequences. The primers begin to use the products from
earlier cycles as template. As a result, the
template/target sequence is copied, as is the primer
that was incorporated into the 5' end of the
20   target/template during its production in the earlier
cycle. The products from these later cycles (which at
the end of the reaction form the vast majority of the
products) consist of target sequence located between the
incorporated primers at the 5' end, and the complements
25   of those primers at the 3' end.
        The method of the invention makes use of chimeric
primers, which are primers which have both a
hybridization segment and a constant segment. The
hybridization segment hybridizes to the template nucleic
30   acid so that extension by the polymerase can occur. The
constant segment does not hybridize with the original
template nucleic acid. As products from earlier cycles
are used as templates, however, this constant segment
also hybridizes to the template. This normalizes the
35   hybridization kinetics across the different target
sequences being simultaneously amplified, preventing

loci from being over or underrepresented at the end of the reaction.

In general, the invention features a method for simultaneously amplifying a plurality of target sequences from template nucleic acid and labeling the amplification products, comprising the steps of (a) combining template nucleic acid and a plurality of pairs of chimeric primers under conditions appropriate for members of the chimeric primer pairs to hybridize to complementary nucleic acid sequences on the template nucleic acid sufficiently well to permit primer extension by a polymerase enzyme, thereby producing template-primer complexes, (b) subjecting the template-primer complexes to conditions appropriate for a first amplification reaction, thereby producing a set of first amplification products, (c) combining these first amplification products and a plurality of pairs of labeled primers, under conditions appropriate for members of the labeled primer pairs to hybridize to complementary nucleic acid sequences on the first amplification products sufficiently well to permit primer extension by a polymerase enzyme, thereby producing product-primer complexes, and (d) subjecting the product-primer complexes to conditions appropriate for a second amplification reaction, thereby producing a set of second labeled amplification products. The template nucleic acid can be isolated nucleic acid, isolated genomic DNA, cDNA, or nucleic acid not isolated away from other cellular components. Each member of a chimeric primer pair includes a hybridization segment and a constant segment. An example of a pair of constant segments is 5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:1) for use on the forward primer and 5'-AATTAACCCTCACTAAAGGGAGA-3' (SEQ ID NO:2) for use on the reverse primer. In the first amplification reaction (step (b)), the conditions include a high concentration of $MgCl_2$ (e.g., from about 2.5 millimolar to about 7.0

millimolar, or about 5 millimolar) and a low extension
temperature (e.g., from about 60°C to about 70°C, from
about 60°C to about 65°C, about 65°C). In the second
amplification reaction (steps (c) and (d)), the labeled
5   primers can be labeled with a biotin molecule, a
fluorophore, a dye, a metal, or a radionuclide. In
addition, the labeled primers have constant segments
identical to the constant segments of the chimeric
primers used in the first amplification reaction.
10   In this embodiment, amplification and detection are
performed in two separate reactions. The first reaction
uses high levels of magnesium and low extension
temperatures, and many (up to several hundred) pairs of
chimeric primers, which are used to simultaneously
15   amplify several hundred target sequences on a template
nucleic acid. Each chimeric primer includes a
hybridization segment, which is nucleic acid which
hybridizes to its complementary sequence on the template
nucleic acid, and a constant segment, which is nucleic
20   acid which does not hybridize to the template nucleic
acid. The hybridization segment is different for each
chimeric primer and each pair of chimeric primers is
designed so that the hybridization segment of each
member of the pair hybridizes to a sequence on the
25   template nucleic acid which flanks a target sequence.
That is, each pair of chimeric primers hybridizes to the
template nucleic acid, in regions which flank the target
(a portion of the template nucleic acid to be
amplified); each member of a chimeric primer pair
30   hybridizes at one end of the target segment and, thus,
together the pair of chimeric primers flank the target
segment. Stated another way, each target sequence is
amplified by its own unique pair of chimeric primers.
For example, ten primer pairs (twenty primers, each with
35   a different hybridization segment), are needed to
amplify ten different target sequences. The
hybridization segments of the primers are therefore used

to amplify the target sequences on the template nucleic
acid.

The first amplification reaction uses the chimeric
primers to make copies of the target sequences. As
5   these copies are made, the polymerase incorporates each
chimeric primer into the copy being made.

A second amplification reaction follows, which
labels the products from the first amplification
reaction. This second reaction uses labeled primers
10 which hybridize to the complements of the constant
segments located at the 3' ends of the amplification
products.

The constant segments of the chimeric primers (and
therefore the labeled primers complementary to the
15 products of those constant segments), are chosen so as
to comprise sequence that is unrelated to the template
nucleic acid, and is therefore unlikely to hybridize
with it. For chimeric primers used to amplify mammalian
DNA, for example, bacteriophage or insect sequences can
20 serve as the constant segments. In one embodiment of
the present invention, two constant segments derived
from bacteriophage are used. One of the pair consists
of the T7 promoter sequence
5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:1), which is
25 synthesized onto the 5' end of the 'forward' chimeric
primer within each primer pair, and the other is the T3
promoter sequence 5'-AATTAACCCTCACTAAAGGGAGA-3' (SEQ ID
NO:2), which is synthesized onto the 5' end of the
'reverse' primer. In this embodiment therefore, two
30 different labeled primers are used. The first is a
biotinylated T7 promoter sequence (5'-biotin-
TAATACGACTCACTATAGGGAGA-3', SEQ ID NO:3), which
hybridizes to the amplification products produced by the
chimeric primer containing the T3 promoter sequence (SEQ
35 ID NO:2). The second labeled primer is a biotinylated
T3 promoter sequence (5'-biotin-AATTAACCCTCACTAAAGGG-3',
SEQ ID NO:4), which hybridizes to the amplification

product produced by the T7 chimeric primer (SEQ ID
NO:1). In other words, a single constant segment is
used on all of the forward primers, and another constant
segment is used on all of the reverse primers. These
5   constant segments normalize hybridization across all of
the different loci, that is, they provide for similar
hybridization kinetics for the many loci being amplified
simultaneously. This results in roughly equal amounts
of products from each locus. Use of the constant
10 segments provide an additional advantage in that they
are used for detection in the second amplification
reaction.

In another embodiment, a *single* sequence is used as
the constant segment on *all* of the chimeric primers
15 (rather than two sequences as in the embodiment above).
In such a case, the 'forward' and the 'reverse' chimeric
primers have identical 5' tails, and a single labeled
primer is, therefore, used to label the amplification
products.
20   In either embodiment, the first amplification
reaction amplifies the sequences on the template nucleic
acid and the second amplification reaction labels the
reaction products from the first amplification reaction.
The combination of chimeric primers and reaction
25 conditions allows the *simultaneous* amplification of
hundreds of loci. For example, as described herein, 558
loci have been successfully amplified with a 50% pass
rate, 92 loci with an 85% pass rate, and 46 or 23 loci
exhibited 90% and 92% pass rates, respectively. An
30 advantage of the invention is that these pass rates were
achieved without further optimization. Alternatively,
the conditions may be further optimized to achieve
higher pass rates.

The invention also features a method for
35 simultaneously amplifying a plurality of target
sequences from template nucleic acid and labeling the
amplification products, including the steps of (a)

combining template nucleic acid and a plurality of pairs
of labeled chimeric primers, under conditions
appropriate for members of the labeled chimeric primer
pairs to hybridize to complementary nucleic acid

5   sequences on the template nucleic acid sufficiently well
to permit primer extension by a polymerase enzyme,
thereby producing template-primer complexes; and (b)
subjecting the template-primer complexes to conditions
appropriate for an amplification reaction, thereby

10 producing a set of labeled amplification products. The
amplification conditions may include a high
concentration of $MgCl_2$ (e.g., from about 2.5 millimolar
to about 7.0 millimolar, about 5 millimolar), and a low
extension temperature (e.g., from about 60°C to about

15 70°C, from about 60°C to about 65°C, about 65°C). The
template nucleic acid can be isolated nucleic acid,
isolated genomic DNA, cDNA, or nucleic acid not isolated
away from other cellular components. Each member of a
labeled chimeric primer pair includes a hybridization

20 segment and a labeled constant segment. For example,
the sequence of the constant segment can be
5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:1) for the
forward primer and 5'-AATTAACCCTCACTAAAGGGAGA-3' (SEQ ID
NO:2) for the reverse primer). The constant segment can

25 be labeled with a biotin molecule, a fluorophore, a dye,
a metal, or a radionuclide.

In another embodiment, amplification and labeling
of the products are carried out in a single step. The
chimeric primers can be synthesized as described above,

30 and labeled (e.g., biotinylated) at the 5' end. In this
embodiment, the reaction conditions are as for the first
amplification reaction, that is, the conditions include
high concentrations of $MgCl_2$ and low extension
temperatures. The second amplification reaction is

35 omitted.

In addition, the invention features a kit for
simultaneously amplifying a plurality of target

sequences from a template nucleic acid and labeling the
amplification products. The kit includes (a) a
plurality of pairs of chimeric primers, and (b) at least
one reaction mixture appropriate for use in
5   amplification reactions. At least one of the reaction
mixtures includes a high concentration of $MgCl_2$ (e.g.,
from about 2.5 millimolar to about 7.0 millimolar, about
5 millimolar). Each member of a chimeric primer pair
includes a hybridization segment and a constant segment
10 (e.g., 5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:1) for
the forward primer and 5'-AATTAACCCTCACTAAAGGGAGA-3'
(SEQ ID NO:2) for the reverse primer). The constant
segment may be labeled (e.g., with a biotin molecule, a
fluorophore, a dye, a metal, or a radionuclide).

15   The invention also relates to kits containing the
chimeric primer pairs and, optionally, reaction mixtures
for practicing the method of the invention. Such kits
can contain a collection of primer pairs useful to
amplify a particular set of target sequences on a
20 template nucleic acid.

        The invention has the advantage of allowing the
simultaneous amplification of many (e.g., several
hundred) target sequences in a single reaction and of
allowing for labeling of the mixture of products. A
25 variety of methods are available for detecting and
analyzing these products. Using size-based methods of
detection (e.g., gels) can be difficult, due to the
large number of different products that are created by
the invention. The Examples below describe analysis by
30 means of genotyping chips, which is a non-size-dependent
method of analysis.

        The kit of the invention also has an advantage in
that it contains a collection of primer pairs chosen so
as to yield a particular type of information. For
35 example, the primer pairs can be chosen to detect
susceptibility to a set of genetic diseases. For use in
forensics studies, the primers can be selected to detect

polymorphisms in target sequences in highly variable
regions of DNA. The polymorphisms found in an
individual's DNA in those variable regions can be
compared to the polymorphisms found in DNA from crime
5   scene evidence in those same regions. If the
polymorphisms from the individual and the evidence are
different, then the individual is excluded from the pool
of possible suspects. If they are the same, the
individual cannot be excluded.

10   The methods and kits of the present invention can
be used in humans and non-humans. For example, the
methods, primers and kits can be used to assay
polymorphisms in animals for veterinary purposes. For
instance, sets of primers can be chosen to amplify
15   target sequences known to be associated with
susceptibilities to diseases with genetic components, or
to detect known genetic defects in purebred animals such
as dogs or horses. Primer sets can also be chosen to
assess levels of biodiversity in populations of animals,
20   plants, or microorganisms.

The methods and kits of the invention can also be
used to amplify sequences across species. For instance,
chimpanzees and humans share approximately 99% sequence
similarity. The methods and kits of the invention can
25   be used to locate those areas in which the 1%
interspecific difference is located, thereby pinpointing
the "evolutionary hotspots" responsible for species
differentiation, and interspecific conserved regions, as
well.

30   Kits can also be created to fingerprint proprietary
biological material. For example, a set of primers can
be chosen corresponding to specific genotypes known to
exist in a protected crop cultivar. Assays of plants
can be made according to the present invention, to
35   determine if those plants correspond to the genotype of
the patented cultivar.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred
5   embodiments of the invention, as illustrated in the accompanying drawings.

Figure 1 is a schematic diagram illustrating the use of chimeric primers in the amplification and labeling reactions, and the products made.
10   Figures 2A and 2B are more detailed diagrams and flow charts illustrating the same process as depicted in Figure 1. Figure 2A depicts the first amplification reaction, and Figure 2B depicts the second amplification reaction (the labeling reaction). "Target" indicates
15 the target sequence, whether on the template nucleic acid, or the amplification products. "zzzz" is meant to indicate nucleic acid sequence outside of the area of interest. In these diagrams, the forward chimeric primer has the sequence 5'-GGGTAAT-3', comprising the
20 hybridization segment "TAAT", and constant segment "GGG". The reverse primer (3'-GCCGTTT-5') has hybridization segment "CGGC" and constant segment "TTT". The labeled primers are 5'-biotin-GGG-3' (which hybridizes to the "CCC" amplified from (and therefore
25 complementary to) the forward chimeric primer), and 3'- TTT-biotin-5' (which hybridizes to the "AAA" amplified from (and therefore complementary to) the reverse chimeric primer).

Figure 3 is a diagram illustrating the relative
30 relationships on the template nucleic acid among the various target sequences, chimeric primers and primer pairs, and labeled primers of the method of the invention. "Target 1", "target 2", etc. denote the target sequences of SNP locus 1, 2, etc. "H1a", "H2a",
35 etc. are the hybridization segments for the forward primers for target 1, target 2, etc. "H1b", "H2b", etc. are the hybridization segments for the reverse primers

for target 1, target 2, etc. "Ca" is the constant
segment for all of the forward chimeric primers, and
"Cb" is the constant segment for all of the reverse
chimeric primers.

5   DETAILED DESCRIPTION OF THE INVENTION
        Described herein is a method of multiplex
amplification. Using this method, several hundred pairs
of chimeric primers are used to simultaneously amplify
several hundred target sequences on template nucleic
10 acid. In one embodiment of the method, two successive
amplification reactions are performed. The first of the
two amplification reactions uses high levels of
magnesium, low extension temperatures, and pairs of
chimeric primers to simultaneously amplify a large
15 number (e.g., up to several hundred) target sequences on
a template nucleic acid. By "high" levels of magnesium
is meant at least 2.5 millimolar, preferably about 2.5
millimolar to about 7.0 millimolar, and more preferably
about 5 millimolar. By a "low" extension temperature is
20 meant, for a polymerase with a normally optimal
extension temperature of 72°C, an extension temperature
of about 60°C to about 70°C, preferably about 60°C to
about 65°C, and most preferably about 65°C. The
amplification enzyme should be subjected to a sub-
25 optimal temperature during extension, so as to equalize
amplification across all of the different primer parts
in the reaction, and to prevent over-representation of
the products of some loci over others. The chimeric
primers used in the method each contain a hybridization
30 segment and a constant segment. The hybridization
segment hybridizes to the template nucleic acid in the
vicinity of a target sequence; the constant segment
consists of a sequence which does not hybridize to the
template.
35   In this embodiment, a second amplification reaction
follows, which uses labeled primers to detectably label

the reaction products of the first amplification
reaction. The combination of chimeric primers and
reaction conditions allows for the *simultaneous*
amplification of a large number of loci. With these
5 methods, for example, 558 loci have been successfully
amplified with a 50% pass rate, 92 loci have been
amplified with an 85% pass rate, and 46 or 23 loci with
90% and 92% pass rates, respectively.

"Amplification" refers broadly to a process for
10 using a polymerase and a pair of primers for increasing
the amount of a particular nucleic acid sequence,
referred to as a target sequence, relative to the amount
of that sequence initially present (on the template
sequence). The process may be accomplished by the *in
15 vitro* methods of the polymerase chain reaction or ligase
chain reaction, or others. A target sequence is a
sequence that lies between the hybridization regions of
two members of a pair of primers and is amplified by
them. The target sequence generally exists as part of a
20 larger "template" sequence; however, in some cases, a
target sequence and the template are the same. The
template sequence may be an isolated nucleic acid or,
alternatively, a nucleic acid which has not been
separated away from the cellular components of the
25 biological source from which it was obtained. Although
"template sequence" generally refers to the nucleic acid
sequence initially present, the products from each
amplification cycle are in fact used as template
sequence in subsequent amplification cycles.

30 The template nucleic acid can be isolated by
methods well known in the art. By "nucleic acid" is
meant a length of DNA, RNA, cDNA, nucleic acids from
mammals or other animals, plants, insects, bacteria,
viruses, or other organisms. Nucleic acids referred to
35 herein as "isolated" are nucleic acids substantially
free of (*i.e.*, separated away from) the cellular
components of the biological source from which they were

obtained (*e.g.*, as it exists in cells or in a mixture of
nucleic acids such as a library), which may have
undergone further processing. "Isolated" nucleic acids
include nucleic acids obtained by methods described

5   herein, similar methods or other suitable methods,
including essentially pure nucleic acids, nucleic acids
produced by chemical synthesis, by combinations of
biological and chemical methods, and recombinantly
produced nucleic acids which are isolated (see *e.g.*,

10 Daugherty, B.L. *et al.*, Nucleic Acids Res. 19(9):2471-
2476 (1991); Lewis, A.P. and Crowe, J.S., Gene 101:297-
302 (1991)). The template nucleic acid can therefore be
a mixture of different nucleic acids or can be lengths
of nucleic acid which are approximately the same (*e.g.*,

15 many copies of the same nucleic acid, such as nucleic
acids which have been chemically synthesized or
recombinantly produced). Alternatively, the template
may occur as nucleic acid which has not been isolated
from other cellular components, but has been treated in

20 such a way that it is available for hybridization with
amplification primers. For example, cells containing
the template DNA can be denatured at high temperature to
break the cellular membranes and expose the nucleic
acids, and the primers and amplification reaction

25 ingredients added directly to the denatured cellular
"soup". Nucleic acids released in this way may be used
in the method of the invention, although the success
rate may not be as high as that reported herein by the
Applicants.

30   A "primer" is a length of single-stranded nucleic
acid, which is used in combination with a polymerase to
amplify a region from a template nucleic acid. Primers
are generally short (*e.g.*, 15-30 bases), but can be
longer if required. The primer must contain a sequence

35 which hybridizes with the template nucleic acid under
the conditions used. Primers may be used singly, that
is, a single primer consisting only of a single sequence

can be used in the amplification reaction, and will
produce one copy of one strand of the template per cycle
of amplification. This can be done in situations where
a large number of copies is not required, or where only
5   one strand is to be copied (e.g., in producing antisense
products), or if the sequence at the other end of the
template is unsuitable for choosing a second primer.

More generally, a pair of primers is used in an
amplification reaction. The two are of different
10 sequences, and are used in combination, and produce a
copy of each template strand per cycle of amplification.
The two different primers should not be complementary to
each other, or they will hybridize to each other rather
than the template, and the polymerase will then be
15 unable to make a copy of the template. Commonly, the
two primers are chosen from sequence at the 5' end of
each of the two complementary strands of the template
nucleic acid.

A "chimeric primer pair" is a set of two chimeric
20 primers wherein the hybridization segments each
hybridize to the template sequence on opposite flanks of
the target sequence. A "chimeric primer" is a primer
(e.g., a short piece of single-stranded nucleic acid)
used for amplification of a target nucleic acid, wherein
25 the primer contains a sequence (i.e., the "hybridization
segment") which hybridizes to the template nucleic acid
in the vicinity of the target sequence, and a sequence
which does not hybridize to the template nucleic acid.
The method of the invention uses chimeric primers
30 comprising a hybridization segment and a constant
segment. The constant segment should not hybridize to
other constant segments, or to the hybridization
segments of the chimeric primers. The constant segment
is generally about 15 to about 35 base pairs, such as
35 about 20 to about 25 base pairs, about 15 to about 25
base pairs, about 15 to about 30 base pairs, about 20 to
about 30 base pairs, about 15 to about 35 base pairs,

about 20 to about 35 base pairs, about 25 to about 35
base pairs, or about 30 to about 35 base pairs. The
hybridization segments of the members of a given
chimeric primer pair should have approximately the same

5   T_m. That is, all of the segments that hybridize with
the template will have melting temperatures within
degrees of each other.

The hybridization segments are selected by methods
commonly known in the art of nucleic acid amplification.

10 That is, the hybridization segment is analogous to, and
serves the same purpose as, the primer as it is commonly
used in the art of nucleic acid amplification, and it is
selected in the same way. To select a pair of primers
(or in the present invention, a pair of hybridization

15 sequences) for amplifying a particular target sequence,
the sequence of the target nucleic acid must be
approximately known (e.g., it has been sequenced in the
organism being studied, or in a related organism). A
short stretch of sequence at either end of the target is

20 then selected to serve as the primer. The two primers
which are intended to amplify a specific target are
chosen on the basis of several characteristics,
including length (e.g., 15-35 base pairs is common in
the field), melting temperature, or other criteria. For

25 instance, depending on the sequence at one end of the
target, it may be necessary to choose a longer or
shorter primer in order to obtain a melting temperature
to match the primer at the other end of the target
sequence. Such selection criteria are known in the

30 field of nucleic acid amplification, and computer
programs are available (e.g., PRIMER (Daly, M.J.,
Lincoln, S.E., and Lander, E.S., unpublished); Lerman,
L.S., and Silverstein, K., Meth. Enzymol. 155:482-501
(1987)) which analyze sequence and choose candidate

35 primers on the basis of specified parameters such as
desired primer length and melting temperature.

The constant segments of the chimeric primer pairs
are not selected to hybridize at the ends of the target
nucleic acid. In fact, the constant segments are
specifically selected so that they *not* hybridize with
5   the template nucleic acid. In one embodiment, the
method of the invention uses two constant segments, the
first comprises the T7 promoter sequence
(5'-TAATACGACTCACTATAGGGAGA-3', SEQ ID NO:1), which is
synthesized onto the 5' end of the 'forward' primer.
10   The second constant segment comprises the T3 promoter
sequence (5'-AATTAACCCTCACTAAAGGGAGA-3', SEQ ID NO:2),
which is synthesized onto the 5' end of the reverse
primer. Although these constant segments were chosen
from T3 and T7 sequences, sequences from other organisms
15   (*e.g.*, insects, reptiles) might also prove useful, so
long as those sequences generally lack the ability to
hybridize with mammalian DNA. On the other hand,
mammalian sequences might be used as the constant
segments in situations where one wishes to construct
20   chimeric primers to amplify non-mammalian template.
There is no requirement that the constant segments have
a complete lack of ability to hybridize to the template
DNA, just that it tends not to hybridize with the
template DNA in general.
25   Creation of an amplification product starts at the
site where the primer hybridizes to the template nucleic
acid, and during extension, the primer itself is
incorporated into the product, resulting in a copy of
the template nucleic acid which differs from the
30   original in that it is truncated at the 5' end where the
primer hybridized and extension began (Figure 1). As
the amplification reaction progresses through its
cycles, the amplification products begin to outnumber,
and later overwhelm, the original template sequences.
35   Primers in later amplification cycles increasingly use
the products from previous cycles as template, so that
at the end of the reaction, the majority of the

amplification products consist of sequences which
terminate at the 5' ends in the primer sequences (Figure
1 and Figures 2A and 2B (boxes)). The 3' ends of the
products terminate with the *complement* of the chimeric
5 primer (Figures 2A and 2B (dashed boxes)). The method
of the invention uses chimeric primers with a non-
hybridizing sequence at the 5' end, resulting in
reaction products terminating in the constant sequences
(Figure 2A, "Products after successive amplification
10 cycles").

The method of the invention, in one of its
embodiments, uses a second amplification reaction to
label the reaction products of the first reaction
(Figure 2A). Labeled primers are used in this reaction.
15 A "labeled primer" is a primer (*e.g.*, a short segment of
single-stranded nucleic acid) which has a detectable
label (*e.g.*, biotin, fluorophore, radioactivity, heavy
metals, dyes, etc.) attached to it. Two primers are
used in this method: 5'-biotin-
20 TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:3), and 5'-biotin-
AATTAACCCTCACTAAAGGG-3' (SEQ ID NO:4). These hybridize
with the constant segments of the chimeric primers, and
the products of this second reaction also have the
primers incorporated into them. These primers are
25 labeled however, and the products are therefore
detectable.

Figure 1 presents a simplified illustration of one
embodiment of the method, which uses two separate
amplification reactions. The first amplifies the
30 targets. The products from this reaction are then
detectably labeled in the second reaction. In Figures
2A and 2B, which illustrate the process in greater
detail, the target is flanked by two sequences ("ATTA"
and "CGGC") which are recognized by the hybridization
35 segments on the chimeric primers ("TAAT" and "GCCG").
The constant segments of the chimeric primers ("GGG" and
"TTT") do not hybridize with the template. The target

is copied from the 3' end of the chimeric primer.
During the next cycle of amplification, this copy of the
target also serves as a template. Unlike the original
template nucleic acid however, this target/template is

5 truncated at the 5' end, ending in the chimeric primer
that was incorporated into the target/template during
the previous cycle of amplification. These
target/templates which are truncated at the 5' ends are
then used as templates in the next cycle. The products

10 of this next cycle are truncated at the 3' end, as well
as the 5' end (Figure 2B).

Figure 3 illustrates the relationships on the
template nucleic acid of the target loci, the chimeric
primers, and the hybridization and constant segments.

15 The template nucleic acid is depicted by the thin
horizontal line, and the target loci (e.g., "target 1",
"target 2", etc.) are symbolized by heavier bars on the
template. "H" denotes a hybridization segment, and "C"
a constant segment. "H1" and "C1" are the hybridization

20 and constant segments for a chimeric primer intended to
amplify "target 1", for example. The suffix "a"
designates that the chimeric primer is the "forward"
primer, and "b" indicates the "reverse" primer. From
this figure, it can be seen that for x target loci,

25 there will be $2x$ different hybridization segments, and
therefore $2x$ different chimeric primers. Those $2x$
different chimeric primers have among them two different
constant segments, however, one for the forward chimeric
primer for each target locus, and another for the

30 reverse.

The method of the invention uses high levels of
magnesium and low extension temperatures for the first
amplification reaction, and more standard conditions for
the labeling reaction.

35 For the first reaction, a 50 μl reaction volume
should contain between approximately 20 and 200 ng of
template nucleic acid (preferably 100 ng), about 0.1 to

about 1.0 μM of each chimeric primer each chimeric primer should be at about 0.1 to 1.0 micromolar concentration, (preferably 0.5 to 1.0 μM), about 1 unit of amplification enzyme, about 0.5 to about 2.0 mM dNTPs

5   (preferably 1 mM), about 10 mM Tris-HCl (pH 8.3), about 5 mM KCl, about 2.5 to about 7.0 mM MgCl$_2$ (preferably 5 mM), and about 0.001% gelatin.

The hybridization segments of the chimeric primers are selected so as to have a T$_m$ (melting temperature) of

10   about 55°C to about 60°C, with a preferred T$_m$ of about 57°C. During the first amplification reaction, an annealing temperature of about 50°C to about 57°C is preferred, and the extension temperature should be about 70°C to about 60°C, preferably about 65°C.

15   In the second amplification reaction, the ingredients are at concentrations commonly used by those of skill in the art, and the temperature conditions are likewise as those commonly used.

In another embodiment of the invention, the second

20   amplification reaction could be omitted if the chimeric primers themselves were labeled (i.e., if each chimeric primer comprised a hybridization segment and a labeled constant segment). If the constant segments of the chimeric primers were labeled (if the 5' ends of the

25   chimeric primers were biotinylated, for example), then the first amplification reaction would simply create labeled amplification products. Extrapolating from the embodiment described above, for example, such a labeled constant segment could comprise the biotinylated T7

30   promoter sequence (5'-biotin-TAATACGACTCACTATAGGGAGA-3', SEQ ID NO:3), added to the 5' end of the hybridization segment of the 'forward' primer. The second constant segment would comprise a labeled T3 promoter sequence (5'-biotin-AATTAACCCTCACTAAAGGGAGA-3', SEQ ID NO:4),

35   added to the 5' end of the reverse primer. To ensure stability of hybridization, it might be necessary to

increase or decrease the length of the constant segment
in some cases.

5   One of skill in the art will recognize that there
exist a variety of detection schemes that can be used
with the Applicants' method, including
chemiluminescence, radioactivity, fluorophores, dyes,
heavy metals, or staining. There are also a wide number
of substitutions that can be made and alternative ways
of practicing the Applicants' method. These
10   alternatives are known to those of skill in the art from
reading the scientific literature and are also available
in compendiums of common laboratory procedures (*e.g.*,
Maniatis, *et al.*, *Molecular Cloning: A Laboratory
Manual*, Cold Spring Harbor Laboratory Press, New York;
15   Ausubel, F.M. *et al.*, eds., *Current Protocols in
Molecular Biology*; Erlich, H.A., ed., *PCR Technology*,
Stockton Press, New York (1989). All publications,
patent applications, patents, and other references
mentioned herein are incorporated by reference in their
20   entirety.

Another method of practicing the invention is to
use a single constant segment, instead of the two
segments described above and illustrated in Figures 2A,
2B and 3. In Figures 2A and 2B, such a change would be
25   shown by all of the constant segments having the same
sequence (*e.g.*, "GGG" *or* "TTT" instead of "GGG" *and*
"TTT"), and in Figure 3, all the constant segments would
simply be labeled "C" (not "Ca" and "Cb").

When multiplexing (amplifying many loci
30   simultaneously), it is imperative to be able to
correctly differentiate between the amplification
products from different loci. There are two commonly-
used methods of accomplishing this; (1) size
discrimination, or (2) on the basis of differentially
35   labeling the reaction products. The first method is
most commonly done with detection methods which separate
on the basis of size (*e.g.*, gels). One chooses loci to

amplify together based on the ability of those loci to create products of different sizes, *e.g.*, locus 1 creates a product of 100 bp, locus 2 creates a 200 bp, etc. Choosing which loci to amplify together is labor-

5   intensive and requires careful planning. It is also possible that new, previously uncharacterized alleles that appear in the "zone" of another locus' reaction products may be misassociated with that other locus. Size-based analysis methods are therefore limited by the

10 number of loci that can be unambiguously separated in the space offered by the separation method (*e.g.*, the length of the gel). The second method, differentially labeling the reaction products, generally involves the attachment of different tags onto the primers for

15 different loci. Fluorescent tags are the most commonly used because one can use molecules which fluoresce at different wavelengths, with the result that amplification products from different loci fluoresce in different colors. This method is limited, however, by

20 the number of different fluorescent molecules available on the market. Both of these methods impose inherent limitations on the number of loci that can be multiplexed together. Another possibility is to use a method modified from that described by Church and

25 Kieffer-Higgins (Science 240:185-188 (1988), wherein the amplification products are separated on a gel, which is then blotted, and successively probed with radiolabelled versions of the amplification primers. This method is also labor-intensive, in addition to being slow.

30   "Genotyping chips" can be used to detect, and differentiate between, the amplification products of the different loci. A "genotyping chip" is a high-density array of oligonucleotide probes to which a sample is hybridized. Hybridization to one of the

35 oligonucleotides indicates that the sample is positive for that oligonucleotide sequence (Fodor, S.P.A. *et al.*, Science 251:767-773 (1991)). If aligned, overlapping

sequences are arrayed (*e.g.*, ABC, BCD, CDE, etc.), then
the chip can be used for sequencing (Pease, A.C. *et al.*,
Proc. Natl. Acad. Sci. USA 91:5022-5026 (1994)).
Alternatively, the chip can contain oligonucleotides
5   corresponding to polymorphic alleles (Chee, M. *et al.*,
Science 274:610-614 (1996)). If there are ten different
alleles at a particular locus, for instance, then
oligonucleotides corresponding to those ten alleles can
be placed on a genotyping chip. Hybridization to one of
10   the alleles indicates that the sample is positive for
that allele (see Example 2, "Design of Genotyping
Chips," *infra*).

In the method of the invention, the hybridization
segments of the chimeric primers are chosen so as to
15   produce relatively short target sequences (*e.g.*, 70-80
base pairs). Amplification products of such similar
lengths would be extremely difficult to differentiate by
size separation methods (*e.g.*, electrophoretic gels),
but are well-suited to detection by genotyping chips, or
20   other size-insensitive detection methods.

The method of the invention also allows for the
simultaneous detection of many polymorphisms in a
sample. A "polymorphism" is, fundamentally, an allelic
variation between the nucleic acids of two samples.
25   Such variations can range from gross morphological
differences to differences in biochemistry, conformation
of biomolecules, or differences in nucleic acid
sequences. The "samples" being examined can be whole
organisms, or portions thereof, and can represent single
30   individuals, or pooled populations. Polymorphisms
include differences in nucleotide sequence, mutations,
insertions, deletions, point mutations, or structural
differences, as well as strand breaks or chemical
modifications that result in an allelic variant in the
35   form of a mismatch. A polymorphism between two nucleic
acids can occur naturally, or be caused by exposure to
or contact with chemicals, enzymes, or other agents, or

can be caused by circumstances which cause damage to
nucleic acids (e.g., exposure to ultraviolet radiation,
mutagens or carcinogens).

The examples below illustrate use of the method of
5   the invention to detect single nucleotide polymorphisms,
or "SNPs," which are polymorphisms that consist of a
difference in a single nucleotide. However, two
sequences being compared may differ at more than a
single nucleotide position. If the second difference
10 occurs within the region where a primer should
hybridize, that primer will not hybridize, and no
reaction will occur. If the two differences both occur
within the region that is amplified, aberrant patterns
will be seen during the detection. In either case, it
15 will be clear that the locus must be investigated
further.

Multiplex amplification of large collections of
mapped SNPs will provide a powerful tool for genetic
studies, some of which are described below.

20 **Linkage Mapping**

Any type of DNA polymorphism can be used to trace
inheritance of disease genes in pedigrees (Lander, E.S.
and Schork, N., Science 265:2037-2048 (1994)). For
technical reasons, geneticists had primarily employed
25 restriction fragment length polymorphisms (RFLPs) during
the 1980s and simple sequence length polymorphisms
(SSLPs or microsatellites) during the 1990s. Both of
these methods assay length differences by gel
electrophoresis. SNPs are now being seen as an
30 attractive alternative for the future, because there
exist a variety of assays that may allow greater
automation, parallelism, and throughput than can be
achieved with length measurement (Conner, B.J. et al.,
Proc. Natl. Acad. Sci. USA 80:278-282 (1983); Cronin,
35 M.T. et al., Hum. Mutation 7:244-255 (1996)). SNPs have
only two alleles and thus are less informative than

typical multi-allelic SSLPs, but this deficiency can be
offset by greater density: a genome scan with 1,000
well-spaced SNPs will extract roughly the same linkage
information as the current standard of 400 well-spaced
5   SSLPs (Kruglyak, L., Nat. Genet. 17:21-24 (1997)). In
addition, as shown below, SNPs can be assayed more
efficiently than SSLPs, thereby providing a true advance
over SSLP methodology.

Linkage Disequilibrium Mapping

10   Linkage disequilibrium mapping extends genetic
analysis from families to populations by using high
density genetic maps to recognize chromosomal segments
descended from a common ancestor (Lander, E.S. &
Botstein, D., Cold Spring Harbor Symp. Quant. Biol.
15   51:49-61 (1986)). Such analyses are useful in
determining the location of a chromosomal segment
conferring disease susceptibility. Ancestral segments
that occur significantly more often in affected
individuals are likely to harbor a susceptibility
20   allele. The precise SNP density required to detect
linkage disequilibrium varies with the age and structure
of the population.

Association Studies

         Beyond simply serving as chromosomal markers, some
25   SNPs are actually the cause of functional variation in a
gene. Several authors have suggested cataloguing
coding-region SNPs (cSNPs) in all 100,000 human genes
and performing association studies between these SNPs
and different phenotypic traits (Risch, N. & Merikangas,
30   K., Science 273:1516-1517 (1996); Lander E.S., Science
274:536-539 (1996); Collins, F.S., Guyer, M.S. and
Chakravarti, A., Science 278:1580-1581 (1997)). For
example, the association between the ApoE gene and
Alzheimer's disease, or between the Factor V gene and
35   deep vein thrombosis could have been discovered in this

fashion (Strittmatter, W.J. and Roses, A.D., Ann. Rev.
Neurosci. 19:53-77 (1996); Voorberg, J., Lancet
343:1535-1536 (1994)).

Other Applications

5   The invention has the advantage of allowing the
simultaneous amplification of many (*e.g.*, several
hundred) target sequences in a single reaction, and of
allowing for labeling of the mixture of products.
         Kits can be made for use with the method of the
10   invention. Such kits would contain pairs of chimeric
primers intended to label specific target sequences on
the template nucleic acid of an organisms or organisms,
and also labeled primer pairs to label the amplification
products. If the chimeric primers in the kit have a
15   single constant segment (rather than two; *i.e.*, one for
the forward primer, and one for the reverse), then only
a single labeled primer need be included. Components
for the two different amplification reactions may also
be included (*e.g.*, amplification reaction buffers,
20   polymerase, etc.). If the amplification and labeling
are done in a single reaction, then reaction components
for only that single reaction need be included.
         While the Examples below demonstrate use of the
methods and kits of the invention in the detection of
25   SNPs, the methods are not limited to this use. The
methods and kits of the invention can be used in any
situation where it is desirable to amplify a large
number of target sequences, regardless of whether or not
they contain a polymorphism. For instance, primers can
30   be chosen to amplify regions thought to be conserved, in
order to rapidly identify individuals polymorphic in
those regions. The methods and kits of the invention
can also be used to evaluate interspecific polymorphism
(such as between humans and chimpanzees) to locate both
35   conserved regions and also "hotspots" of evolutionary
change.

The method and kits of the invention can be used
for forensic identification of individuals. Primer
pairs can be selected to detect polymorphisms in target
sequences in highly variable regions of DNA. The
5   polymorphisms found in a suspect's DNA in those variable
regions can be compared to the polymorphisms found in
DNA from crime scene evidence in those same regions.
Matching polymorphisms includes the suspect in the pool
of possible perpetrators, while differing polymorphisms
10 excludes the suspect. These same sets of primers can
also be useful in "biometrics" applications, which is
the identification and verification of individuals via a
unique biological profile. Potential applications of
biometrics include the identification of deceased
15 persons, verification of identities of prisoners slated
for release, or verification of a person's access to
sensitive information or areas.

The kit of the invention has an advantage in that
it can be assembled to contain a collection of primer
20 pairs chosen so as to yield a particular type of
information. For example, the primer pairs can be
chosen to detect susceptibility to a genetic disease or
set of diseases, or the presence of pathogens or
parasites. Chromosomal deletions known to lead to
25 cancer or other diseases can also be detected in this
manner.

The methods and kits of the present invention can
be used in humans and non-humans. For example, the
methods, primers and kits can be used to assay sequences
30 in animals for veterinary purposes (e.g., presence of
pathogens or parasites). Sets of primers can be chosen
to amplify target sequences known to be associated with
susceptibilities to diseases with genetic components, or
to detect known genetic defects in purebred animals such
35 as dogs or horses.

The primer sets in the kits can also be chosen to
assess levels of biodiversity in field populations of

animals, plants, or microorganisms. Individual
organisms can also be "fingerprinted" and later re-
identified, e.g., in animal migration studies, for
instance. These same kits can also be used to study the
5   evolution and migration of animal and plant populations.

Kits can also be created to fingerprint proprietary
biological material, such as microbiological strains,
crop cultivars, or animals. For example, a set of
primers can be chosen corresponding to highly variable
10 regions known to exist in a protected crop cultivar.
Assays of plants can be made according to the present
invention, to create unique genetic profiles, and to
determine if those plants correspond to the genotype of
the patented cultivar. Parentage of purebred animals
15 can also be verified in this way.

The invention is further described in the following
examples, which are intended to illustrate, and not
limit, the scope of the invention described therein and
in the claims.

20 **EXAMPLES**
Methods
Individuals Surveyed for SNPs
The individuals surveyed from chosen from CEPH
pedigrees K104, K884 and K1331, from the Amish,
25 Venezuelan and Utah populations, respectively. The SNP
survey by gel-based sequencing examined three unrelated
individuals (K104-1, K884-2, K1331-1) and a pool of ten
individuals (K104-13, -14, -15, -16; K884-15, -16;
K1331-12, -13, -14, -15).

30 **DNA Sequencing of SNPs from Sequence Tagged Sites (STSs)**
STSs were amplified with their corresponding
amplification primers as described in Hudson, T.J. et
al. (Science 270:1945-1954 (1995)) and Dietrich et al.
(Dietrich, W. F. et al., Nature 380:149-152 (1996);
35 Dietrich, W. F. et al., Nature Genetics 7:220-245;

Dietrich, W. et al., Genetics 131:423-447 (1992)), which
are herein incorporated by reference in their entirety.
The forward primer was modified to include the M13 -21
primer site 5'-TGTAAAACGACGGCCAGT-3' (SEQ ID NO:5) at

5   its 5'-end. The resulting amplification products were
subjected to dye-primer sequencing (Koontz, W.L.G. and
Fukunaga, K., IEEE Trans. Comp., C-21,171-178 (1972),
herein incorporated by reference in their entirety),
with products detected on an ABI377 or ABI373

10   fluorescence sequence detector. Possible sequence
variations were detected by the ABI *Sequence Navigator*
software package, which suggests potential heterozygotes
by identifying nucleotide positions at which a secondary
peak exceeds a selected threshold of 50%. Such apparent

15   variations were then visually inspected, to compare the
patterns seen among the several individuals.

Example 1:   Obtaining SNPs
A: SNPs from Sequence Tagged Sites (STSs)
          SNPs were obtained by surveying sequence-tagged

20   sites (STSs) distributed across the human genome. These
STSs are short genomic sequences, each with a
corresponding amplification assay. An initial
collection of 1,139 STSs was chosen from among the
24,568 STSs that had been used in the construction of a

25   physical map of the human genome at the Whitehead/MIT
Center for Genome Research (Hudson, T.J. et al., Science
270:1945-1954 (1995) and Schuler, G. et al., Science
274:540-546 (1996), herein incorporated by reference in
their entirety). These STSs contain a total of 279 kb

30   of genomic sequence between all of the amplification
primer sites, with one-third consisting of random
genomic sequence and two-thirds of 3'-ends of expressed
sequence tags (3'-ESTs) and primarily representing
untranslated regions of genes.

35   Each STS was amplified from four samples: three of
the samples were individual human DNAs and the fourth

was pooled DNA from ten individuals. The amplification
products were subjected to single-pass DNA sequencing
using fluorescent-dye primers and gel electrophoresis;
sequence traces were compared by a computer program
5. followed by visual inspection (See "Methods," *supra*).
Candidate SNPs were declared when two alleles were seen
among the three individuals, with both alleles present
at a frequency greater than 30% in the pooled sample.
The term 'candidate SNP' is used because a subset of
10 such apparent polymorphisms turn out to be sequencing
artifacts, as discussed below in Example 3, *infra*.

The survey found 279 candidate SNPs, corresponding
to a rate of one SNP per 1001 base pairs screened and an
observed nucleotide heterozygosity of 3.96×10^{-4} (Table
15 1). The SNPs were distributed among 239 STSs, with some
containing more than one polymorphic site. The
polymorphism rate was lower in 3'-ESTs than in random
genomic sequence, although the difference fell just
short of statistical significance ($p = 0.057$,
20 one-sided). The ratio of transitions to transversions
was 2:1. Although the dinucleotide CpG comprises only
about 2% of the sequence surveyed, nearly 25% of the
SNPs occurred at such sites with the substitution nearly
always being C<->T. Cytosine residues within CpG
25 dinucleotides are well known to be the most mutable
sites within the human genome, because most are
methylated and can spontaneously deaminate to yield a
thymidine residue (Cooper, D.N. & Karwczak, M., Hum.
Genet. 85:55-74 (1990)). In addition to the single base
30 substitutions, 23 insertion/deletion polymorphisms were
also found (with all but eight involving a single base),
corresponding to a frequency of one per 12 kb surveyed.

-31-

Table 1:   Results of SNP screening

	Gel-based Sequencing			Chip-Based Detection		
	All STSs	3'-ESTs	random genomic STSs	All STSs	3'-ESTs	random genomic STSs
No. STSs screened	1,139	705	434	16,725	12,649	4,076
Total bases screened	279,165	186,524	92,641	1,981,030	1,324,320	656,710
No. candidate SNPs found	279	161	118	2,748	1,749	999
SNP frequency	1/1001	1/1159	1/785	1/721	1/757	1/657
Heterozygosity $(x10^{-4})$	3.96 ± 0.38	3.42 ± 0.43	5.04 ± 0.67	4.58 ± 0.15	4.36 ± 0.18	5.02 ± 0.28
No. STSs containing SNPs	239	137	102	2,299	1,515	784
% transitions among SNPs	67%	67%	67%	70%	70%	71%
% SNPs occurring within CpG	24%	23%	25%	24%	25%	22%

B: SNPs from other sources

Additional SNPs were isolated by using two directed approaches using public databases. First, reports in the literature of common variants in gene coding regions were collected. Out of 143 cases tested, 121 variants were confirmed by the detection of two alleles in the four-sample screening panel described above. The remaining 22 reported variants may be true polymorphisms, but simply monomorphic in the individuals tested. The second approach utilized the GenBank database, which contains multiple entries for some ESTS. Such entries were compared to identify single-nucleotide differences, which might reflect either common polymorphisms or sequencing errors in single-pass EST

sequencing. 200 such apparent differences were tested and the presence of a SNP was confirmed in 93 cases. These two directed approaches thus yielded an additional 214 SNPs.

5 **Example 2: Design of Genotyping Chips**

Gel-based sequencing was satisfactory for the initial screen (Example 1, above), but a more streamlined approach is needed in a large-scale survey. High-density DNA probe arrays provide an alternative

10 approach for analyzing DNA sequences (Chee, M. *et al.*, Science 274:610-614 (1996); Kozal, M.J. *et al.*, Nat. Medicine 2:753-759 (1996)). Such 'DNA chips' can be made by using parallel light-directed chemistry to synthesize specified oligonucleotide probes covalently

15 bound at defined locations on a glass surface or 'chip' (Fodor, S.P.A. *et al.*, Science 251:767-773 (1991); Pease, A.C. *et al.*, Proc. Natl. Acad. Sci. USA 91:5022-5026 (1994)). Current technology allows fabrication of 1.28 cm x 1.28 cm arrays of ~320,000

20 distinct oligonucleotides, each residing in a 'feature' of ~20 x 25 microns and containing more than 10^7 copies of the oligonucleotide.

A target DNA sequence of length L can be screened for polymorphism by hybridizing a biotin-labeled sample

25 to a variant detector array (VDA) of size 8 x L. For each position on both strands, the array has four 25-mer probes complementary to the sequence centered at the position. The four differ only in that the central (13th) position is substituted by each of the four

30 nucleotides. Individuals that are homozygous for the expected sequence (*e.g.*, "A-A") should hybridize more strongly to the perfectly complementary probe than to the three probes containing a central mismatch. The presence of an SNP would be expected to give rise to a

35 different hybridization pattern, with homozygotes showing strong hybridization to an alternative base

(e.g., "G-G") and heterozygotes (e.g., "A-G") showing
hybridization to two probes (e.g., "A" and "G"). The
VDA thus signals the presence of a sequence variant and,
in many cases, indicates the nature of the change.

5   VDAs have been used previously for mutation
detection of small, well-studied DNA targets in large
numbers of samples, including 387 base pairs from the
HIV-1 genome, 3.5 kb from the BRCA1 gene, and 16.6 kb
from the human mitochondrion (Chee, M. et al., Science
10   274:610-614 (1996); Kozal, M.J. et al., Nat. Medicine
2:753-759 (1996); Hacia, J.G. et al., Nat. Genet.
14:441-447 (1996)). In this setting, the normal
hybridization pattern can be characterized with great
precision and mutations detected with high accuracy.

15   The current project, however, sought to use VDAs in
a large-scale survey. A total of 16,725 STSs covering 2
Mb of human DNA was selected, one-third from random
genomic sequence and two-thirds from 3'-ESTs. The
survey employed 149 distinct chip designs, each
20   containing 150,000-300,000 features.

The STSs were examined in seven unrelated
individuals (K104-1, -16, K884-2, -15, -16, K1331-12, -
13), chosen from the CEPH pedigrees K104 (Amish), K884
(Venezuelan), and K1331 (Utah). Altogether, this
25   represented a survey of about 14 Mb of genomic sequence.
For each chip, the corresponding STSs were amplified
from an individual, pooled together, biotin-labeled,
hybridized and stained as follows.

STSs were amplified with their corresponding
30   primers as described in Hudson et al. (Science
270:1945-1954 (1995)). Amplification products intended
for hybridization to the same chip (typically
amplification products from 100-200 STSs from a single
individual) were pooled together for subsequent
35   processing. Approximately 1-2 μg of the pooled
amplification product was purified with Qiaquick
purification kit (Qiagen, Hilden, Germany), fragmented

with DNAse I, then biotin-labeled with terminal
deoxynucleotidyl transferase (TdT, Life Technologies,
GibcoBRL, (Gaithersburg, Maryland), according to the
manufacturer's instructions. The fragmentation was

5   performed in a 40 μl reaction with 0.2 unit of DNase I
(Promega, Madison, Wisconsin), 10 mM Tris-acetate (pH
7.5), 10 mM magnesium acetate, and 50 mM potassium
acetate at 37°C for 15 minutes, after which the reaction
was stopped by heat inactivation at 96°C for 15 minutes.

10   The terminal transferase reaction was performed by
adding 15 units of TdT and 12.5 mM biotin-N6-ddATP
(DuPont-NEN Products, Boston, MA) to the preceding
reaction mixture, incubating at 37°C for 1 hour and then
heat-inactivating at 96°C for 15 minutes.

15   The labeled samples were hybridized to the chip as
follows. Samples were denatured at 96°C for 5-6 minutes
and cooled on ice for 2-5 minutes. Chips were
pre-hybridized with 6X SSPET (0.9 M NaCl, 60 mM NaH_2PO_4,
6 mM EDTA (pH 7.4), 0.005% Triton X-100) for

20   approximately 5 minutes and then hybridized with the
denatured sample in hybridization buffer (3M
tetramethylammonium chloride, 10 mM Tris-HCl (pH 7.8), 1
mM EDTA, 0.01% Triton X-100, 100 mg/ml herring sperm
DNA, and 200 pM control oligonucleotide) at 44°C for 15

25   hours on a rotisserie at 40 RPM. Chips were washed 3
times with 1X SSPET and 10 times with 6X SSPET at 22°C,
then stained at room temperature with staining solution
(2 mg/ml streptavidin R-phycoerythrin (Molecular Probes,
Eugene, Oregon) and 0.5 mg/ml acetylated BSA in 6X

30   SSPET) for 8 minutes. After staining, chips were washed
10 times with 6X SSPET at 22°C on a fluidics workstation
(Affymetrix, Santa Clara, California). Hybridization to
the chip was detected by using a confocal chip scanner
(HP/Affymetrix, Santa Clara, California) with a

35   resolution of 40-80 pixels per feature and a 560 nm
filter, and visual inspection.

At each position, samples were classified as homozygous for the expected sequence, homozygous for an alternative sequence, or heterozygous. A collection of 2,748 candidate SNPs was identified, corresponding to a

5   rate of one per 721 base pairs surveyed and an observed nucleotide heterozygosity of 4.58×10^{-4} (Table 1, supra). The number of STSs containing SNPs was 2,299. The SNPs had a mean heterozygosity of 33%, with the minor allele having a mean frequency of 25%. SNPs were

10 found less often in 3'-ESTs than in random genomic sequence ($p < 0.023$, one-sided).

The nucleotide heterozygosity rate was indistinguishable from the estimate obtained from gel-based sequencing in Example 1, above ($p > 0.12$,

15 two-sided test). The ratio of transitions to transversions and the proportion of SNPs occurring at CPG dinucleotides were also indistinguishable. The frequency of SNPs was higher in the chip-based survey because more samples (14 vs. 6 haploid genomes) were

20 surveyed.

Example 3:   Evaluation of Accuracy of Gel-based vs. Chip-based Surveys

Large-scale surveys are generally imperfect, and it is therefore important to assess the error rates of the

25 two systems. False positive rates were estimated by retesting candidate SNPs by thorough multipass gel-based sequencing. In the single-pass gel-based sequencing, 16% of the 120 candidate SNPs proved to be false positives, compared to 12% of the 220 candidate SNPs

30 found in the chip-based survey. False negative rates were estimated by including a subset of STSs in both surveys. Fifty-five SNPs were identified with both of the two survey methods, and were carefully confirmed to eliminate false positives. Eight (15%) of the 55 were

35 missed by single-pass gel-based resequencing and seven (13%) missed by the chip-based survey. Many of the

errors were due to random factors, because they were
eliminated simply by repeating the original experiment.
However, some errors were reproducible artifacts that
could be eliminated only by changing the detection

5 protocol (for example, by using dye-terminators rather
than dye-primers in gel-based sequencing). The rates
were broadly similar in the two surveys, corresponding
to roughly one false positive and false negative every
6,000-10,000 bases.

10   The error rates reflect the particular
implementation of the large-scale survey (single-pass
coverage for gel-based sequencing; one-color
hybridization at one temperature to a single VDA
design), rather than inherent limitations of the

15 technologies. It is likely that at the expense of
additional effort, both technologies can provide higher
accuracy still in future large-scale screens.

With current limitations, candidate SNPs should be
confirmed before being regarded as certain. We

20 initially confirmed SNPs by using gel-based
resequencing, but subsequently developed an accurate
chip-based method using genotyping chips (Example 4,
below).

Example 4: Genotyping Chips for Multiplex Amplification

25   Efficient methods are required for large-scale
genotyping of SNPs, and one approach is to extend the
use of chip-based re-sequencing from SNP discovery to
SNP genotyping (Cronin, M.T. et al., Hum. Mutation
7:244-255 (1996)).

30   Genotyping chips containing 'genotyping arrays'
were synthesized for each SNP to be tested. Each
genotyping array consists of two short VDAs
corresponding to the two alternative alleles. The
presence of an allele is reflected in strong

35 hybridization to the corresponding resequencing array.
Amplification assays were designed for the region

containing each SNP, with the goal of being robust and
mutually compatible. This was done by ensuring that (1)
the amplification targets were small (typically a few
nucleotides around the polymorphic site), (2) the
5   primers all had similar calculated melting temperatures,
and (3) constant sequences were added to the 5'-ends of
the forward and reverse primers to facilitate
batch-labeling of pooled amplification products. Each
assay was tested to ensure that it amplified a single
10   fragment from genomic DNA.

For each SNP, primers were chosen using the PRIMER
software package (Hudson, T.J. et al., Science
270:1945-1954 (1995)) to closely flank the polymorphic
base and to have a predicted melting temperature of
15   57°C. Forward and reverse primers were synthesized with
the T7 (5'-TAATACGACTCACTATAGGGAGA-3', SEQ ID NO:1) and
T3 (5'-AATTAACCCTCACTAAAGGGAGA-3', SEQ ID NO:2) promoter
sites at their respective 5'-ends. Each primer pair was
individually tested to determine if it produced a single
20   clear fragment visible by agarose gel electrophoresis
and ethidium bromide staining, as described in Hudson et
al. (Science 270:1945-1954 (1995)). Amplification
assays passing this test were further classified as
being "strong" or "weak," according to the yield of the
25   fragment produced. Primer pairs were grouped into
multiplex sets, with the sets chosen to consist of
either all strong assays or all weak assays.

The most complex genotyping chip tested contained
genotyping arrays for 558 candidate SNPs identified in
30   the chip-based survey. Initially, the 558 loci were
separately amplified, then pooled, labeled and
hybridized to the chip. To determine whether each locus
could be reliably read, a formal 'detection test' was
defined: a locus passed if, for each of three
35   individuals tested, the expected DNA sequence could be
successfully read on both strands for one or both
alleles. In all, 98% of the loci passed this detection

test, with the remaining ten failing due to weak
hybridization or cross-hybridization.

Example 5: Multiplex Amplification

In multiplex amplification, primer pairs from many
5   different loci are combined in a single reaction and
amplified simultaneously. Specifically, multiplex
amplification reactions were performed in a 50 μl volume
containing 100 ng of human genomic DNA, 0.5-1.0 μM of
each primer, 1 unit of AmpliTaq Gold (Perkin-Elmer,
10 Foster City, California), 1 mM dNTPs, 10mM Tris/HCl (pH
8.3), 50 mM KCl, 5 mM $MgCl_2$, and 0.001% gelatin.
Thermocycling was performed on a Tetrad (MJ Research,
Watertown, Massachusetts), with initial denaturation at
96°C for 10 minutes followed by 30 cycles of
15 denaturation at 96°C for 30 seconds, primer annealing at
55°C for 2 minutes, and primer extension at 65°C for 2
minutes. After 30 cycles, a final extension reaction
was carried out at 65°C for 5 minutes. Because the
resulting amplification products were small, it was
20 unnecessary to fragment them (as was done for the STSs
in the SNP screen in Example 2, *supra*). The products
were then biotin-labeled in a standard amplification
reaction, by using T7 and T3 primers with biotin-labels
at their 5'-ends (5'-biotin-TAATACGACTCACTATAGGGAGA-3'
25 (SEQ ID NO:3), and 5'-biotin-AATTAACCCTCACTAAAGGG-3'
(SEQ ID NO:4), respectively). The reaction was
performed with 1 μl of template DNA, 0.5-1.0 μM of
labeled primer, 1 unit of AmpliTaq Gold (Perkin-Elmer,
Foster City, California), 100 μM dNTPs, 10mM Tris-HCl
30 (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, and 0.001% Gelatin.
Thermocycling was performed with initial denaturation at
96°C for 10 minutes followed by 25 cycles of
denaturation at 96°C for 30 seconds, primer annealing at
52°C for 1 minute, and primer extension at 72°C for 1
35 minute. After 25 cycles, a final extension reaction was
carried out at 72°C for 5 minutes. The amplification

products from the various multiplex reactions for an
individual were then pooled together. One-tenth of the
pooled sample was denatured and used for chip
hybridization. Chips were hybridized, washed, stained
5 and scanned, as described in Example 2, *supra*.

Although it is typically difficult to combine many
thermocycling assays, this approach worked surprisingly
well: of the 558 loci, 92% passed the detection test
when amplification was performed in 24 sets of ~23 loci;
10 90% passed when amplified in 12 sets of ~46 loci; 85%
passed when amplified in 6 sets of ~92 loci; and 50%
passed when amplified in a single set of 558 loci. The
success appears to have resulted from a combination of
factors, including the small size of the amplification
15 targets, optimization of amplification conditions and
the presence of the constant sequence at the 5'-ends of
the primers. Unsuccessful assays can be salvaged by
grouping them into additional multiplex sets or by
redesigning the assays.
20 Multiplex amplification of sets of 46 loci were
used in subsequent experiments, because the number of
reactions was decreased by 46-fold while allowing the
vast majority of loci (512 out of 558) to be assayed.
The procedure was further tested in 39 individuals and
25 proved quite consistent: 96% of the 572 loci could be
successfully read in 100% of individuals tested. The
remaining 4% of the loci were successfully read in
nearly all of the individuals.

EQUIVALENTS
30 While this invention has been particularly shown
and described with references to preferred embodiments
thereof, it will be understood by those skilled in the
art that various changes in form and details may be made
therein without departing from the spirit and scope of
35 the invention as defined by the appended claims. Those
skilled in the art will recognize or be able to

ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described specifically herein. Such equivalents are intended to be encompassed in the scope

5   of the claims.

CLAIMS

What is claimed is:

1.   A method for simultaneously amplifying a plurality
     of target sequences from template nucleic acid and
5   labeling the amplification products, the method
     comprising:
     (a) combining template nucleic acid and a
          plurality of pairs of chimeric primers,
          wherein each member of a chimeric primer pair
10   comprises a hybridization segment and a
          constant segment, under conditions appropriate
          for members of the chimeric primer pairs to
          hybridize to complementary nucleic acid
          sequences on the template nucleic acid
15   sufficiently well to permit primer extension
          by a polymerase enzyme, thereby producing
          template-primer complexes;
     (b) subjecting the template-primer complexes to
          conditions appropriate for a first
20   amplification reaction, thereby producing a
          set of first amplification products, with the
          proviso that the conditions include a high
          concentration of $MgCl_2$ and a low extension
          temperature, whereby the target sequences in
25   the template nucleic acid are amplified;
     (c) combining the first amplification products and
          a plurality of pairs of labeled primers,
          wherein each member of a labeled primer pair
          comprises a constant segment, under conditions
30   appropriate for members of the labeled primer
          pairs to hybridize to complementary nucleic
          acid sequences on the first amplification
          products sufficiently well to permit primer
          extension by a polymerase enzyme, thereby
35   producing product-primer complexes; and

(d)   subjecting the product-primer complexes to
conditions appropriate for a second
amplification reaction, thereby producing a
set of second amplification products which are
labeled.

2.   The method of claim 1, wherein the constant segment
for the forward chimeric primer is
5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:1) and the
constant segment for the reverse chimeric primer is
5'-AATTAACCCTCACTAAAGGGAGA-3' (SEQ ID NO:2).

3.   The method of claim 1, wherein the labeled primer
pairs are labeled with a biotin molecule, a
fluorophore, a dye, a metal, or a radionuclide.

4.   The method of claim 1, wherein the $MgCl_2$
concentration is from about 2.5 millimolar to about
7.0 millimolar.

5.   The method of claim 4, wherein the $MgCl_2$
concentration is about 5 millimolar.

6.   The method of claim 1, wherein the extension
temperature is from about 60°C to about 70°C.

7.   The method of claim 6, wherein the extension
temperature is from about 60°C to about 65°C.

8.   The method of claim 7, wherein the extension
temperature is about 65°C.

9.   The method of claim 1, wherein the template nucleic
acid is isolated nucleic acid.

10.   The method of claim 9, wherein the template nucleic
acid is isolated genomic DNA.

11. The method of claim 1, wherein the template nucleic acid is cDNA.

12. The method of claim 1, wherein the template nucleic acid is not isolated away from other cellular
5   components.

13. A method for simultaneously amplifying a plurality of target sequences from template nucleic acid and labeling the amplification products, the method comprising:
10   (a)   combining template nucleic acid and a plurality of pairs of labeled chimeric primers, wherein each member of a labeled chimeric primer pair comprises a hybridization segment and a labeled constant segment, under
15   conditions appropriate for members of the labeled chimeric primer pairs to hybridize to complementary nucleic acid sequences on the template nucleic acid sufficiently well to permit primer extension by a polymerase
20   enzyme, thereby producing template-primer complexes; and
       (b)   subjecting the template-primer complexes to conditions appropriate for an amplification reaction, thereby producing a set of labeled
25   amplification products.

14. The method of claim 13, wherein the constant segment for the forward chimeric primer is 5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:1) and the constant segment for the reverse chimeric primer is
30   5'-AATTAACCCTCACTAAAGGGAGA-3' (SEQ ID NO:2).

15.   The method of claim 13, wherein the labeled primer
      pairs are labeled with a biotin molecule, a
      fluorophore, a dye, a metal, or a radionuclide.

16.   The method of claim 13, wherein the amplification
      conditions include a high concentration of $MgCl_2$.

17.   The method of claim 16, wherein the $MgCl_2$
      concentration is from about 2.5 millimolar to about
      7.0 millimolar.

18.   The method of claim 17, wherein the $MgCl_2$
      concentration is about 5 millimolar.

19.   The method of claim 13, wherein the first
      amplification conditions include a low extension
      temperature.

20.   The method of claim 19, wherein the extension
      temperature is from about 60°C to about 70°C.

21.   The method of claim 20, wherein the extension
      temperature is from about 60°C to about 65°C.

22.   The method of claim 21, wherein the extension
      temperature is about 65°C.

23.   The method of claim 13, wherein the template
      nucleic acid is isolated nucleic acid.

24.   The method of claim 23, wherein the template
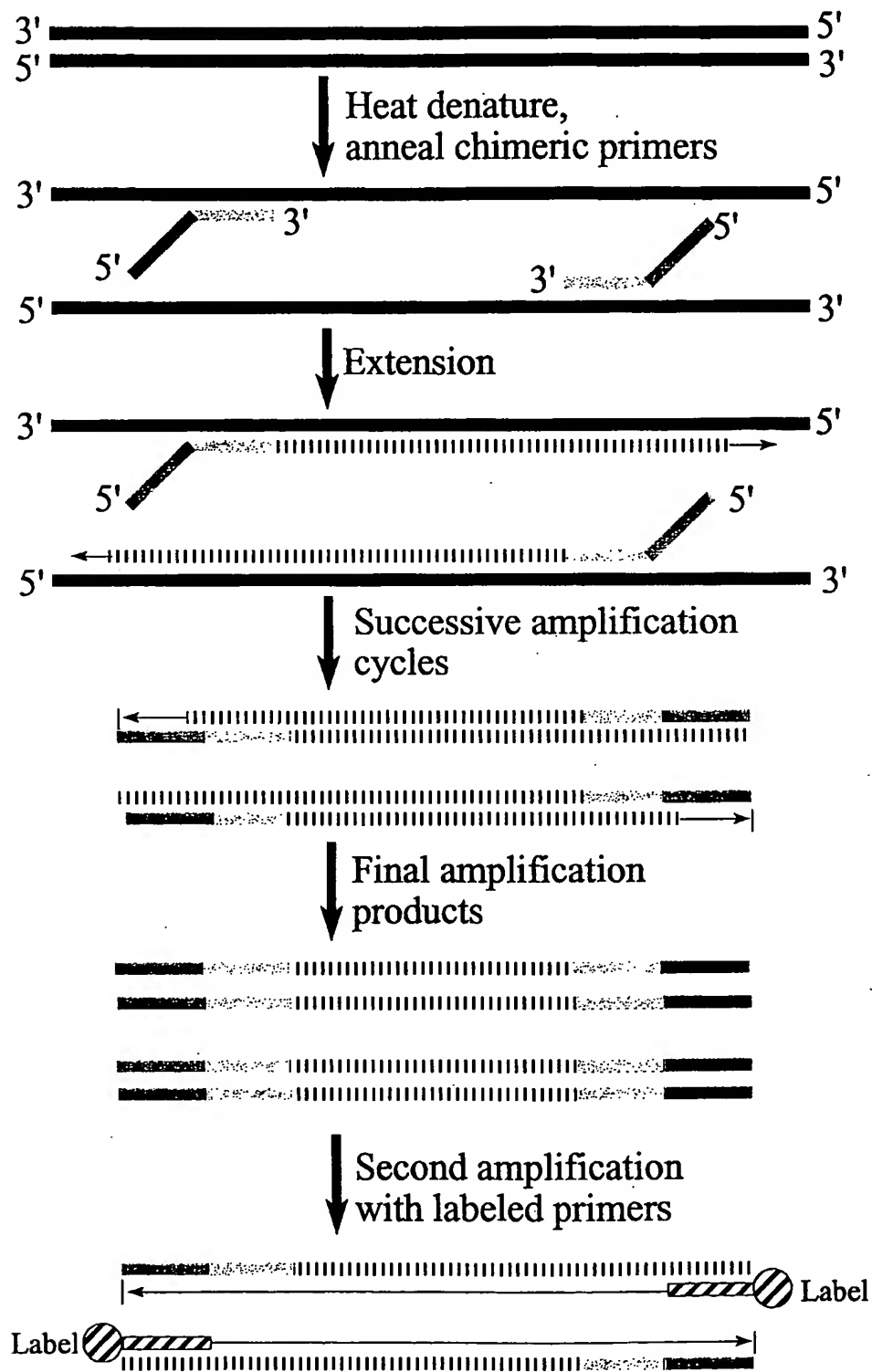   nucleic acid is isolated genomic DNA.

25.   The method of claim 13, wherein the template
      nucleic acid is cDNA.

26. The method of claim 13, wherein the template nucleic acid is not isolated away from other cellular components.

27. A kit for simultaneously amplifying a plurality of target sequences from a template nucleic acid and labeling the amplification products, the kit comprising

   (a) a plurality of pairs of chimeric primers, wherein each member of a chimeric primer pair comprises a hybridization segment and a constant segment, so that under appropriate conditions the members of the chimeric primer pairs can hybridize to complementary nucleic acid sequences on a template nucleic acid sufficiently well to permit primer extension by a polymerase enzyme; and

   (b) at least one reaction mixture appropriate for use in amplification reactions, with the proviso that at least one of the reaction mixtures includes a high concentration of $MgCl_2$;

   so that when template nucleic acid is combined with the chimeric primers of (a) and a reaction mixture of (b) under the appropriate conditions, a plurality of target sequences are amplified from the template nucleic acid.

28. The kit of claim 27, wherein the constant segment for the forward chimeric primer is 5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO:1) and the constant segment for the reverse chimeric primer is 5'-AATTAACCCTCACTAAAGGGAGA-3' (SEQ ID NO:2).

29. The kit of claim 27, wherein the constant segment is labeled.

30. The kit of claim 29, wherein the label is a biotin molecule, a fluorophore, a dye, a metal, or a radionuclide.

31. The kit of claim 27, wherein a reaction mixture with a high $MgCl_2$ concentration has an $MgCl_2$ concentration of about 2.5 to about 7.0 millimolar.

32. The kit of claim 31, wherein a reaction mixture with a high $MgCl_2$ concentration has an $MgCl_2$ concentration of about 5 millimolar.

Fig. 1

3'   5'
5'   3'

↓ Heat denature,
  anneal chimeric primers

3'   5'
                    3'
5'
              3'   5'
5'   3'

↓ Extension

3'   5'

5'
              5'
5'   3'

↓ Successive amplification cycles

↓ Final amplification products

↓ Second amplification with labeled primers

Label   Label

Fig. 2A

FIRST AMPLIFICATION REACTION:

Original Template:

```
(SEQ ID NO: 6)    <-- 3'-zzzz-ATTA-(target)-GCCG-zzzz-5'  -->
(SEQ ID NO: 7)    <-- 5'-zzzz-TAAT-(target)-CGGC-zzzz-3'  -->
```
                    │ Denature template,
                    │ anneal chimeric primers
                    ▼

Initial Amplification Cycle:

```
(SEQ ID NO: 6)    <-- 3'-zzzz-ATTA-(target)-GCCG-zzzz-5'  -->
                      5'-GGG-TAAT-3' ───────────────►
                              ◄─────────── 3'-GCCG-TTT-5'
(SEQ ID NO: 7)    <-- 5'-zzzz-TAAT-(target)-CGGC-zzzz-3'  -->
```
                    │ Products after first
                    │ amplification cycle
                    ▼

```
(SEQ ID NO: 8)    <-- 3'-zzzz-ATTA-(target)-CGGC-TTT-5'

(SEQ ID NO: 9)        5'-GGG-TAAT-(target)-CGGC-zzzz-3'  -->
```
                    │ Denature template,
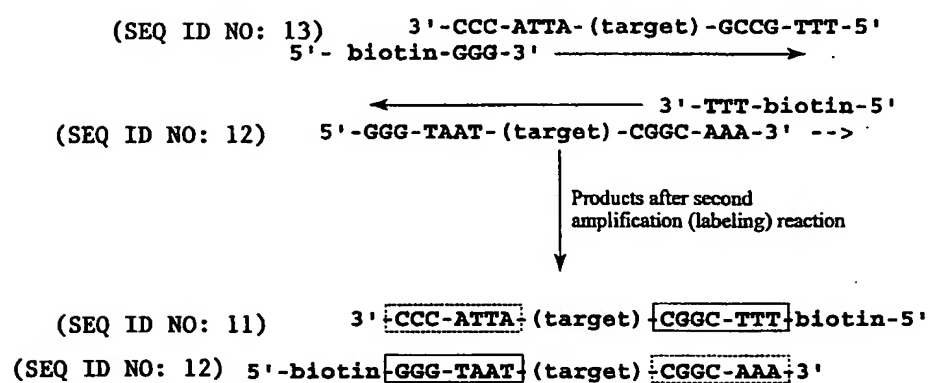   │ anneal chimeric primers
                    ▼

Second Amplification Cycle:

```
(SEQ ID NO: 10)   <-- 3'-zzzz-ATTA-(target)-GCCG-TTT-5'
                      5'-GGG-TAAT-3' ───────────────►

(SEQ ID NO: 9)                ◄─────────── 3'-GCCG-TTT-5'
                      5'-GGG-TAAT-(target)-CGGC-zzzz-3'  -->
```
                    │ Products after successive
                    │ amplification cycles
                    ▼

```
(SEQ ID NO: 11)    3'-CCC-ATTA-(target)-CGGC-TTT-5'

(SEQ ID NO: 12)    5'-GGG-TAAT-(target)-CGGC-AAA-3'
```
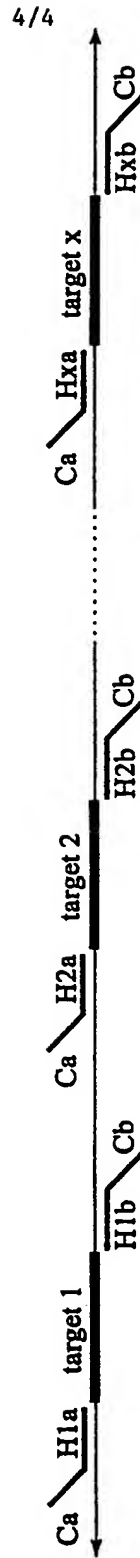
Fig. 2B

SECOND AMPLIFICATION REACTION:

(SEQ ID NO: 13)   3'-CCC-ATTA-(target)-GCCG-TTT-5'
                  5'- biotin-GGG-3' ————————————>

                        <———————————— 3'-TTT-biotin-5'
(SEQ ID NO: 12)   5'-GGG-TAAT-(target)-CGGC-AAA-3' -->

                                   │ Products after second
                                   │ amplification (labeling) reaction
                                   ▼

(SEQ ID NO: 11)   3' CCC-ATTA (target) CGGC-TTT biotin-5'

(SEQ ID NO: 12) 5'-biotin GGG-TAAT (target) CGGC-AAA 3'

Fig. 3